基于线性模型平均估计的置信区间

王苗苗[123]

(1.北京中医药大学中药学院, 北京 100102; 2.中国科学院大学数学科学学院, 北京 100049; 3.中国科学院数学与系统科学研究院, 北京 100190)

摘要 已有的关于模型平均估计渐近分布理论的研究多是基于局部误设定的假设, [1] 是其中开创性的且最著名的文章之一. 虽然利用局部误设定的假设可以证明模型平均估计渐近分布理论, 但是 [2] 等对此假设提出了不合理性质疑和解释. 本文我们研究 [1] 中的置信区间估计方法. 证明了在一般参数设定下, 虽然 [1] 中的渐近分布理论不一定成立, 但是关于不确定参数的线性函数的置信区间在正态分布误差、线性回归模型下是有效的, 即置信区间的覆盖率趋于预设定的名义水平. 我们通过模拟研究进一步验证了理论结果.

关键词 模型平均, 渐近分布, 线性回归.

MR(2000)主题分类号 62E20, 62F12

Confidence Interval based on Model Average Estimator for Linear Regression

WANG Miaomiao1[123]

(1. *School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing* 100102;
2.*School of Mathematical Sciences, University of Chinese Academy of Sciences, , Beijing* 100049;
3.*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190)

Abstract Most authors have examined the inference of model averaging estimators under local misspecification assumption. [1] is one of the most groundbreaking and famous articles. However, although the local misspecification assumption provides a suitable framework for studying the asymptotic theories of Frequentist Model Averaging estimators, it also draws comments from [2] because of its realism. In this paper, we prove that the confidence interval construction method for the linear function of parameters in [1] is still valid in linear regression with normally distributed error under general fixed parameter setup. It means that the coverage probability of the confidence interval can reach the nominal level. The simulation results support our theory conclusion.

Keywords Model averaging, Asymptotic distribution, Linear regression.

1 引 言

模型平均方法自提出以来, 备受统计学家关注, 广泛应用于金融学、经济学、社会学、生物医学等各领域. 模型平均方法有两大研究方向: 贝叶斯模型平均 (Bayes

Model Averaging, BMA) 和频率模型平均 (Frequentist Model Averaging, FMA). 与贝叶斯模型平均 [3] 不同, FMA 中的权重仅由数据确定, 并且不需要先验假设. 本文主要关注 FMA 的研究. 模型平均估计的渐近分布理论是 FMA 方法研究中最核心的问题之一, 关于此的研究非常多. 例如 [1] 和 [4] 研究了基于极大似然估计的模型平均方法的渐近理论. [5] 推导出了线性回归模型中最小二乘模型平均估计的渐近分布. 其他工作还有 [6]、[7]、[8]、[9] 以及 [10] 等. 这些工作都是基于对参数的局部误设定的假设. 此假设要求一些真实参数的阶数和样本量有关, 即为 $1/\sqrt{n}$. 虽然这个假设为研究FMA估计量的渐近理论提供了一个合适的框架, 但实际上此假设有待商榷, 具体评论参阅 [2]. 最近, 在一般固定参数设定下, 也有关于模型平均估计渐近分布理论的研究. 例如, 基于 Mallows 准则和交叉验证准则, 在嵌套候选模型设置下, [11] 推导了最小二乘模型平均估计的渐近分布；去除局部误设定的假设, [12] 研究了模型平均估计的极限分布理论, 但是极限分布中含有候选模型的偏差.

[1] 是众多研究模型平均渐近分布理论的重要文献之一, 给出了基于极大似然估计的模型平均方法的渐近理论. 设独立随机样本 $y_1, \ldots, y_n$, 其中 n 是样本量, 且 y_i 的密度函数为 f. 我们对参数 $\mu = \mu(f)$ 进行统计推断, 这里 μ 是任意光滑函数. [1] 关注如下模型

$$f(y, \boldsymbol{\theta}),$$

其中 $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^T = \left(\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T\right)^T \subset \boldsymbol{\Theta} \subset \mathbb{R}^q$, 这里 $\boldsymbol{\beta}$ 和 $\boldsymbol{\gamma}$ 分别是 q_1 维和 q_2 维参数向量, 且 $\boldsymbol{\beta}$ 是确定在模型中的参数, 而 $\boldsymbol{\gamma}$ 则不然. 若 $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$ 固定且已知, 对应的模型我们称作零模型, 即 $f(y, \boldsymbol{\theta}) = f\left(y, \left(\boldsymbol{\beta}^T, \boldsymbol{\gamma}_0^T\right)^T\right)$. 他们基于一种局部误设定的假定研究模型平均估计的渐近分布. 局部误设定假设真实模型为

$$f_{true}(y) = f_n(y) = f\left(y, \left(\boldsymbol{\beta}_0^T, \left(\boldsymbol{\gamma}_0 + \boldsymbol{\delta}/\sqrt{n}\right)^T\right)^T\right),$$

其中 $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{q_2})^T$ 表示模型偏离零模型的程度, 此时,

$$\mu_{true} = \mu\left(\left(\boldsymbol{\beta}_0^T, \left(\boldsymbol{\gamma}_0 + \boldsymbol{\delta}/\sqrt{n}\right)^T\right)^T\right).$$

局部误设定假设意味着模型中一些参数的阶数和样本量有关, 为 $1/\sqrt{n}$. 尽管在此假设下, 我们可以推导出 FMA 估计的渐近分布理论, 但很多学者对此提出质疑, 例如[2] 怀疑此假设的真实性. 我们研究估计的有效性通常是基于大样本理论, 然而随着样本增多, 若参数阶数为 $1/\sqrt{n}$, 则参数空间变得越来越小, 导致参数越来越接近真实值. 换句话说, 此时所有的候选模型越来越接近真实模型. 所以, 在大样本下, 局部误设定的假设是不真实的. 总之, 这一假设并不是一个常规的条件, 而是关乎理论推导过程及结果的关键假设.

本文在一般固定参数设置下, 研究 [1] 的置信区间构造方法. 我们发现虽然 [1] 中的渐近分布理论不一定成立, 但是关于参数 $\boldsymbol{\gamma}$ 线性函数的置信区间在正态分布误差和线性回归模型下是有效的, 即置信区间的覆盖率可达到预设定的名义水平, 为模型平均估计的区间预测提供了方法和理论支持.

本文具体安排如下：第 2 节主要介绍候选模型的构造方法和模型平均估计. 在第 3 节中, 我们说明在固定参数设置下, [1] 的渐近分布理论不一定成立, 并给出此

设置下 [1] 中的置信区间构造方法. 第 4 节在正态分布误差、线性回归模型下, 推导出模型平均估计的一些理论性质, 并基于这些理论给出不确定参数的线性函数的置信区间. 第 5 节通过模拟研究, 进一步验证第 4 节的置信区间的有效性. 最后, 第 6 节做出总结. 本文中所有引理和定理的证明见附录.

2 模型框架和模型平均估计

考虑基于有限项协变量的线性回归模型

$$y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{\gamma} + e_i, \quad i = 1, \ldots, n, \tag{2.1}$$

其中 y_i 是响应变量, $\boldsymbol{x}_i = (x_{i,1}, \cdots, x_{i,q_1})^T$ 和 $\boldsymbol{z}_i = (z_{i,1}, \cdots, z_{i,q_2})^T$, $i = 1, \ldots, n$ 是协变量向量, 误差项 e_i, $i = 1, \ldots, n$ 独立同正态分布. 令 $\boldsymbol{h}_i = \left(\boldsymbol{x}_i^T, \boldsymbol{z}_i^T\right)^T$ 以及 $H = (\boldsymbol{h}_1, \cdots, \boldsymbol{h}_n)^T$. 设 $\mathrm{E}(e_i|\boldsymbol{h}_i) = 0$ 且 $\mathrm{Var}(e_i|\boldsymbol{h}_i) = \sigma^2$. 记 $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T\right)^T$, 则 y_i 的条件密度函数 $f(y_i, \boldsymbol{\theta}|\boldsymbol{h}_i)$ 同 $\mathcal{N}\left(\boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{\gamma}, \sigma^2\right)$ 的密度函数.

令 $\boldsymbol{\theta}_0$ 表示真实参数值. 设一般固定参数设置为

$$\boldsymbol{\theta}_0 = \left(\boldsymbol{\beta}_0^T, \boldsymbol{\gamma}^T\right)^T = \left(\boldsymbol{\beta}_0^T, (\boldsymbol{\gamma}_0 + \boldsymbol{\delta}')^T\right)^T = \left(\boldsymbol{\beta}_0^T, (\boldsymbol{\gamma}_0 + \boldsymbol{\delta}/\sqrt{n})^T\right)^T, \tag{2.2}$$

其中 $\boldsymbol{\gamma}_0 = (\gamma_{0,1}, \cdots, \gamma_{0,q_2})^T$. 从而 $\boldsymbol{\delta}' = \left(\delta_1', \ldots, \delta_{q_2}'\right)^T = \boldsymbol{\gamma} - \boldsymbol{\gamma}_0$ 是固定的且 $\boldsymbol{\delta} = \sqrt{n}\boldsymbol{\delta}'$.

首先, 在一般固定参数的假设 (2.2) 下, 我们给出候选模型的构造方法. 设有 M 个子模型, 其中第 m 个候选模型包含参数 $\boldsymbol{\beta}$ 以及参数 $\boldsymbol{\gamma}$ 的某些分量 γ_j 等于 $\gamma_{j,0}$. 记 k_m 为第 m 个候选模型中参数的个数. 通常, 如果考虑参数的所有可能的组合, 则 $M = 2^{q_2}$; 如果仅考虑嵌套组合, 则 $M = q_2 + 1$. 由参数下标组成的集合 S 是 $\{1, \ldots, q_2\}$ 的子集. 对第 m 个子模型, 参数下标集$S_m = \{i_1, \cdots, i_{k_m - q_1}\} \subset \{1, \ldots, q_2\}$ 且 $S_m^c = \{i_{k_m - q_1 + 1}, \cdots, i_{q_2}\}$ 是 S_m 在集合 $\{1, \ldots, q_2\}$ 中的补集. 为便于说明, 记 $\boldsymbol{\gamma}_{S_m} = \left(\gamma_{i_1}, \cdots, \gamma_{i_{k_m - q_1}}\right)^T$ 和 $\boldsymbol{\gamma}_{S_m^c} = \left(\gamma_{i_{k_m - q_1 + 1}}, \cdots, \gamma_{i_{q_2}}\right)^T$. 从而, 第 m 个模型中的参数为 $\boldsymbol{\theta}_m = \left(\boldsymbol{\beta}^T, \boldsymbol{\gamma}_m^T\right)^T$, 其中 $\boldsymbol{\gamma}_m = (\gamma_1, \gamma_2, \cdots, \gamma_{q_2})^T$. 此时, $\boldsymbol{\gamma}_{S_m^c} = \boldsymbol{\gamma}_{0,S_m^c}$. 此外, 通过对 $\boldsymbol{\theta}_m$ 的各分量进行一系列的置换, 重写成如下形式

$$\Pi_m\boldsymbol{\theta}_m = \left(\boldsymbol{\beta}^T, \boldsymbol{\gamma}_{S_m}^T, \boldsymbol{\gamma}_{0,S_m^c}^T\right)^T = \left(\boldsymbol{\theta}_{S_m}^T, \boldsymbol{\gamma}_{0,S_m^c}^T\right)^T,$$

其中 $\boldsymbol{\theta}_{S_m} = \left(\boldsymbol{\beta}^T, \boldsymbol{\gamma}_{S_m}^T\right)^T$ 且 Π_m 为置换矩阵. Π_m 可以通过对单位矩阵进行一系列行置换得到. 设单位矩阵 $\boldsymbol{I}_q = (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_q)^T$, 其中 $\boldsymbol{e}_j$ 是第 j 个分量为 1 的单位向量, 则

$$\Pi_m = \left(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_{q_1}, \boldsymbol{e}_{q_1+i_1}, \cdots, \boldsymbol{e}_{q_1+i_{k_m-q_1}}, \boldsymbol{e}_{q_1+i_{k_m-q_1+1}}, \cdots, \boldsymbol{e}_{q_1+i_{q_2}}\right)^T.$$

在第 m 个模型下, 记 $\widehat{\boldsymbol{\theta}}_{S_m}$ 为参数 $\boldsymbol{\theta}_{S_m}$ 在参数空间 $\boldsymbol{\Theta}_m \subset \mathbb{R}^{k_m}$ 上的极大似然估计, 这可由对数似然方程 $\sum_{i=1}^n \partial \log f(y_i, \boldsymbol{\theta}_m|\boldsymbol{h}_i)/\partial\boldsymbol{\theta}_{S_m} = 0$ 解得. 此时, $\widehat{\boldsymbol{\theta}}_{S_m}$ 也是 $\boldsymbol{\theta}_{S_m}$ 的最小二乘估计. 定义 $\widehat{\boldsymbol{\theta}}_m = \Pi_m^T\left(\widehat{\boldsymbol{\theta}}_{S_m}^T, \boldsymbol{\gamma}_{0,S_m^c}^T\right)^T$. 记 $D_n = \widehat{\boldsymbol{\delta}}_{full} = \sqrt{n}\left(\widehat{\boldsymbol{\gamma}}_{full} - \boldsymbol{\gamma}_0\right)$ 且 $\widehat{\boldsymbol{\gamma}}_{full}$ 是基于全模型 $\boldsymbol{\gamma}$ 的估计. 定义 $\widehat{\mu}_m = \mu\left(\widehat{\boldsymbol{\theta}}_m\right)$ 为第 m 个子模型的估计. 下面我们考虑与 [1] 中形式相同的模型平均估计. $\mu(\boldsymbol{\theta}_0)$ 的估计具有如下形式:

$$\widehat{\mu}\left(\boldsymbol{w}\left(D_n\right)\right) = \sum_{m=1}^M w_m\left(D_n\right)\widehat{\mu}_m, \tag{2.3}$$

其中 $w_m(D_n)$ 是对第 m 个候选模型所加的随机权重. 于是权重向量

$$\boldsymbol{w}(D_n) = (w_1(D_n), \ldots, w_M(D_n))^T$$

属于权重集合 $\mathcal{W} = \{\boldsymbol{w}(D_n) \in [0,1]^M : \sum_{m=1}^{M} w_m(D_n) = 1\}$.

3 区间估计

为便于描述 [1] 中的统计推断问题, 我们首先给出一些记号. 设 $f(y,\boldsymbol{\theta})$ 关于参数 $\boldsymbol{\theta}$ 的一阶偏导存在且 $\Psi(y,\boldsymbol{\theta}) = \partial \log f(y,\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, 则 Fisher 信息阵定义为

$$\mathcal{F}(\boldsymbol{\theta}) = E_0\left\{\Psi(y,\boldsymbol{\theta})\Psi(y,\boldsymbol{\theta})^T\right\}.$$

从而, 基于全模型在 $(\boldsymbol{\beta}_0^T, \boldsymbol{\gamma}_0^T)^T$ 处的 Fisher 信息阵为

$$J_{full} = \mathcal{F}\left((\boldsymbol{\beta}_0^T, \boldsymbol{\gamma}_0^T)^T\right) = \begin{array}{c} \begin{array}{cc} q_1\times q_1 & q_1\times q_2 \end{array} \\ \begin{pmatrix} J_{00} & J_{01} \\ & \\ J_{10} & J_{11} \end{pmatrix} \\ \begin{array}{cc} q_2\times q_1 & q_2\times q_2 \end{array} \end{array},$$

且其逆矩阵为

$$J_{full}^{-1} = \begin{array}{c} \begin{array}{cc} q_1\times q_1 & q_1\times q_2 \end{array} \\ \begin{pmatrix} J^{00} & J^{01} \\ & \\ J^{10} & J^{11} \end{pmatrix} \\ \begin{array}{cc} q_2\times q_1 & q_2\times q_2 \end{array} \end{array}.$$

类似于 [1], 我们感兴趣的参数 $\mu(\boldsymbol{\theta}_0)$ 的 $1-\alpha$ 置信区间为

$$\begin{aligned} &\left[\widehat{\mu}(\boldsymbol{w}(D_n)) - \widehat{\boldsymbol{\omega}}^T\left\{D_n - \widehat{\delta}(D_n)\right\}/\sqrt{n} - z_{1-\alpha/2}\widehat{\varphi}/\sqrt{n},\right. \\ &\left.\widehat{\mu}(\boldsymbol{w}(D_n)) - \widehat{\boldsymbol{\omega}}^T\left\{D_n - \widehat{\delta}(D_n)\right\}/\sqrt{n} + z_{1-\alpha/2}\widehat{\varphi}/\sqrt{n}\right], \end{aligned} \tag{3.4}$$

其中 $z_{1-\alpha/2}$ 是标准正态分布的 $1-\alpha/2$ 分位数; $\widehat{J}_{00}^{-1}$, $\widehat{J}_{10}$, $\widehat{J}_{01}$ 以及 $\widehat{J}_{11}$ 分别是 J_{00}^{-1}, J_{10}, J_{01} 以及 J_{11} 的相合估计, 从而利用 $\widehat{K} = \left(\widehat{J}_{11} - \widehat{J}_{10}\widehat{J}_{00}^{-1}\widehat{J}_{01}\right)^{-1}$ 估计 $K = \left(J_{11} - J_{10}J_{00}^{-1}J_{01}\right)^{-1}$; $\widehat{\boldsymbol{\omega}} = \widehat{J}_{10}\widehat{J}_{00}^{-1}\frac{\partial\mu}{\partial\boldsymbol{\beta}} - \frac{\partial\mu}{\partial\boldsymbol{\gamma}}$ 和 $\widehat{\varphi} = \left\{\left(\frac{\partial\mu}{\partial\boldsymbol{\theta}}\right)^T \widehat{J}_{00}^{-1}\frac{\partial\mu}{\partial\boldsymbol{\theta}} + \widehat{\boldsymbol{\omega}}^T\widehat{K}\widehat{\boldsymbol{\omega}}\right\}^{1/2}$ 分别是 $\boldsymbol{\omega} = J_{10}J_{00}^{-1}\frac{\partial\mu}{\partial\boldsymbol{\beta}} - \frac{\partial\mu}{\partial\boldsymbol{\gamma}}$ 和 $\varphi = \left\{\left(\frac{\partial\mu}{\partial\boldsymbol{\theta}}\right)^T J_{00}^{-1}\frac{\partial\mu}{\partial\boldsymbol{\theta}} + \boldsymbol{\omega}^T K\boldsymbol{\omega}\right\}^{1/2}$ 的相合估计, 注意其中的偏导数均是在$(\boldsymbol{\beta}_0^T, \boldsymbol{\gamma}_0^T)^T$ 点取得; Γ_m 是 $(k_m - q_1)\times q_2$ 维选择矩阵, 将向量 $\boldsymbol{\gamma}_m$ 映射成它的子向量 $\boldsymbol{\gamma}_{S_m} = \Gamma_m\boldsymbol{\gamma}_m$; 记 $K_m = \left(\Gamma_m K^{-1}\Gamma_m^T\right)^{-1}$, 令

$$\widehat{\delta}(D_n) = K^{1/2}\left\{\sum_{m=1}^{M} w_m(D_n)V_m\right\}K^{-1/2}D_n,$$

这里 $D_n = \widehat{\boldsymbol{\delta}}_{full} = \sqrt{n}\,(\widehat{\boldsymbol{\gamma}}_{full} - \boldsymbol{\gamma}_0)$ 且 $V_m = K^{-1/2}\Gamma_m^T K_m \Gamma_m K^{-1/2}$.

基于第 m 个模型的 Fisher 信息阵为

$$\begin{array}{cc} & \begin{array}{cc} q_1 \times q_1 & q_1 \times (k_m - q_1) \end{array} \\ J_{S_m} = & \left(\begin{array}{cc} J_{00} & J_{01}\Gamma_m^T \\ & \\ \Gamma_m J_{10} & \Gamma_m J_{11}\Gamma_m^T \end{array} \right), \\ & \begin{array}{cc} (k_m - q_1) \times q_1 & (k_m - q_1) \times (k_m - q_1) \end{array} \end{array}$$

且其逆矩阵为

$$\begin{array}{cc} & \begin{array}{cc} q_1 \times q_1 & q_1 \times (k_m - q_1) \end{array} \\ J_{S_m}^{-1} = & \left(\begin{array}{cc} J^{00,m} & J^{01,m} \\ & \\ J^{10,m} & J^{11,m} \end{array} \right). \\ & \begin{array}{cc} (k_m - q_1) \times q_1 & (k_m - q_1) \times (k_m - q_1) \end{array} \end{array}$$

下面, 我们解释 [1] 中的结果在固定参数设置下不一定成立. 记 $\widehat{\boldsymbol{\theta}}_{S_m} = \left(\widehat{\boldsymbol{\beta}}_{S_m}^T, \widehat{\boldsymbol{\gamma}}_{S_m}^T\right)^T$. 由 $\sum_{i=1}^n \partial \log f\left(y_i, \left(\widehat{\boldsymbol{\theta}}_{S_m}^T, \boldsymbol{\gamma}_{0,S_m^c}^T\right)^T\right) \Big/ \partial \boldsymbol{\theta}_{S_m} = 0$, 将

$$\sum_{i=1}^n \frac{\partial \log f\left(y_i, \left(\widehat{\boldsymbol{\theta}}_{S_m}^T, \boldsymbol{\gamma}_{0,S_m^c}^T\right)^T\right)}{\partial \boldsymbol{\theta}_{S_m}}$$

在 $(\boldsymbol{\beta}_0^T, \boldsymbol{\gamma}_0^T)^T$ 处 Taylor 展开, 得

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial \log f\left(y_i, \left(\widehat{\boldsymbol{\theta}}_{S_m}^T, \boldsymbol{\gamma}_{0,S_m^c}^T\right)^T\right)}{\partial \boldsymbol{\theta}_{S_m}} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\frac{\partial \log f\left(y_i, (\boldsymbol{\beta}_0^T, \boldsymbol{\gamma}_0^T)^T\right)}{\partial \boldsymbol{\theta}_{S_m}} + \left(\frac{\partial^2 \log f\left(y_i, \left(\widetilde{\boldsymbol{\beta}}_0^T, \widetilde{\boldsymbol{\gamma}}_0^T\right)^T\right)}{\partial \boldsymbol{\theta}_{S_m} \partial \boldsymbol{\theta}_{S_m}^T}, \frac{\partial^2 \log f\left(y_i, \left(\widetilde{\boldsymbol{\beta}}_0^T, \widetilde{\boldsymbol{\gamma}}_0^T\right)^T\right)}{\partial \boldsymbol{\theta}_{S_m} \partial \boldsymbol{\gamma}_{S_m^c}^T} \right) \right. \\ &\quad \left. \cdot \left(\left(\widehat{\boldsymbol{\beta}}_{S_m} - \boldsymbol{\beta}_0\right)^T, (\widehat{\boldsymbol{\gamma}}_{S_m} - \boldsymbol{\gamma}_{0,S_m})^T, \mathbf{0}^T \right)^T \right] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\frac{\partial \log f\left(y_i, (\boldsymbol{\beta}_0^T, \boldsymbol{\gamma}_0^T)^T\right)}{\partial \boldsymbol{\theta}_{S_m}} + \frac{\partial^2 \log f\left(y_i, \left(\widetilde{\boldsymbol{\beta}}_0^T, \widetilde{\boldsymbol{\gamma}}_0^T\right)^T\right)}{\partial \boldsymbol{\theta}_{S_m} \partial \boldsymbol{\theta}_{S_m}^T} \right. \\ &\quad \left. \cdot \left(\left(\widehat{\boldsymbol{\beta}}_{S_m} - \boldsymbol{\beta}_0\right)^T, (\widehat{\boldsymbol{\gamma}}_{S_m} - \boldsymbol{\gamma}_{0,S_m})^T \right)^T \right], \end{aligned}$$

其中 $\left(\widetilde{\boldsymbol{\beta}}_0^T,\widetilde{\boldsymbol{\gamma}}_0^T\right)^T=\left(\widetilde{\boldsymbol{\beta}}_0^T,\widetilde{\boldsymbol{\gamma}}_{0,S_m}^T,\boldsymbol{\gamma}_{0,S_m^c}^T\right)^T$ 位于 $\left(\boldsymbol{\beta}_0^T,\boldsymbol{\gamma}_0^T\right)^T$ 和 $\left(\widehat{\boldsymbol{\theta}}_{S_m}^T,\boldsymbol{\gamma}_{0,S_m^c}^T\right)^T$ 之间. 因此

$$\begin{pmatrix}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{S_m}-\boldsymbol{\beta}_0\right)\\ \sqrt{n}\left(\widehat{\boldsymbol{\gamma}}_{S_m}-\boldsymbol{\gamma}_{0,S_m}\right)\end{pmatrix}=\left(-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f\left(y_i,\left(\widetilde{\boldsymbol{\beta}}_0^T,\widetilde{\boldsymbol{\gamma}}_0^T\right)^T\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T}\right)^{-1}\cdot\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\log f\left(y_i,\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}_{S_m}}\right].$$

根据 [13] 中定理 18 的证明, 注意到 $\mathrm{E}_0\left\{\Psi_m(y,\boldsymbol{\theta}_{S_m})\right\}$ 关于 $\boldsymbol{\theta}_{S_m}$ 连续, 可知任给 $\varepsilon>0$, 存在 $\rho>0$ 使得当 $|\boldsymbol{\theta}_{S_m}-\boldsymbol{\theta}_{0,S_m}|<\rho$ 时,

$$\left|\mathrm{E}_0\dot{\Psi}_m(y,\boldsymbol{\theta}_{S_m})+J_{S_m}\right|<\varepsilon. \tag{3.5}$$

再由一致强大数律知, 依概率 1 地存在正整数 N, 使得当 $n>N$ 时有

$$\sup_{\boldsymbol{\theta}_{S_m}\in\{\boldsymbol{\theta}_{S_m}:|\boldsymbol{\theta}_{S_m}-\boldsymbol{\theta}_{0,S_m}|<\rho\}}\left|\frac{1}{n}\sum_{i=1}^{n}\dot{\Psi}_m(y_i,\boldsymbol{\theta}_{S_m})-\mathrm{E}_0\dot{\Psi}_m(y,\boldsymbol{\theta}_{S_m})\right|<\varepsilon. \tag{3.6}$$

结合 (3.5) 以及 (3.6), 当 $n>N$ 时, 有

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f\left(y_i,\left(\boldsymbol{\beta}_0^T,\boldsymbol{\gamma}_0^T\right)^T\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T}+J_{S_m}\right|\\ \leq&\sup_{\boldsymbol{\theta}_{S_m}\in\{\boldsymbol{\theta}_{S_m}:|\boldsymbol{\theta}_{S_m}-\boldsymbol{\theta}_{0,S_m}|<\rho\}}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\dot{\Psi}_m(y,\boldsymbol{\theta}_{S_m})-\mathrm{E}_0\dot{\Psi}_m(y,\boldsymbol{\theta}_{S_m})\right|\right.\\ &\left.+\left|\mathrm{E}_0\dot{\Psi}_m(y,\boldsymbol{\theta}_{S_m})+J_{S_m}\right|\right\}\leq 2\varepsilon.\end{aligned}$$

但是在固定参数设置下, **不能**保证 $\left(\widehat{\boldsymbol{\theta}}_{S_m}^T,\boldsymbol{\gamma}_{0,S_m^c}^T\right)^T\xrightarrow{a.s.}\left(\boldsymbol{\beta}_0^T,\boldsymbol{\gamma}_0^T\right)^T$. 从而, 我们**不能**保证下式成立

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f\left(y_i,\left(\widetilde{\boldsymbol{\beta}}_0^T,\widetilde{\boldsymbol{\gamma}}_0^T\right)^T\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T}\xrightarrow{a.s.}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f\left(y_i,\left(\boldsymbol{\beta}_0^T,\boldsymbol{\gamma}_0^T\right)^T\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T}.$$

注意到

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f\left(y_i,\left(\widetilde{\boldsymbol{\beta}}_0^T,\widetilde{\boldsymbol{\gamma}}_0^T\right)^T\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T}+J_{S_m}\right|\\ =&\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f\left(y_i,\left(\widetilde{\boldsymbol{\beta}}_0^T,\widetilde{\boldsymbol{\gamma}}_0^T\right)^T\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T}-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f\left(y_i,\left(\boldsymbol{\beta}_0^T,\boldsymbol{\gamma}_0^T\right)^T\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T}\right.\\ &\left.+\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f\left(y_i,\left(\boldsymbol{\beta}_0^T,\boldsymbol{\gamma}_0^T\right)^T\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T}+J_{S_m}\right|,\end{aligned}$$

所以**不能**保证

$$-\sum_{i=1}^{n}\partial^2\log f\left(y_i,\left(\widetilde{\boldsymbol{\beta}}_0^T,\widetilde{\boldsymbol{\gamma}}_0^T\right)^T\right)\Big/\left(n\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T\right)\xrightarrow{a.s.}J_{S_m}.$$

对 $f(y_i,\boldsymbol{\theta}_0)$ 在$(\boldsymbol{\beta}_0^T,\boldsymbol{\gamma}_0^T)^T$ 处 Taylor 展开知,

$$f(y_i,\boldsymbol{\theta}_0)=f\left(y_i,(\boldsymbol{\beta}_0^T,\boldsymbol{\gamma}_0^T)^T\right)\left\{1+\frac{\partial \log f\left(y_i,(\boldsymbol{\beta}_0^T,\boldsymbol{\gamma}_0^T)^T\right)}{\partial \boldsymbol{\gamma}}\boldsymbol{\delta}'+R_1(y_i,\boldsymbol{\delta}')\right\},$$

其中 $R_1(y_i,\boldsymbol{\delta}')$ 是余项. 然而, 在一般固定参数设置下, $f(y_i,\boldsymbol{\theta}_0)$ 的余项不一定随着 n 趋于无穷而趋于 0. 故而, [1] 的渐近分布理论不一定成立.

4 主要结论

本节在模型框架 (2.1) 和固定参数设置 (2.2) 下, 我们证明当 μ 是参数 $\boldsymbol{\gamma}$ 的线性函数时, (3.4) 的覆盖率可达到预设定的名义水平. 为便于清晰解释 [1] 中置信区间构造方法在特殊设置下的有效性, 下面, 我们仍采用类似于上节的记号. 在给定 H 下, 我们给出信息阵的具体形式. 基于全模型在 $(\boldsymbol{\beta}_0^T,\boldsymbol{\gamma}_0^T)^T$ 处的 Fisher 信息矩阵为

$$J_{full|H}=\begin{pmatrix}J_{00|H} & J_{01|H}\\ J_{10|H} & J_{11|H}\end{pmatrix}=\frac{1}{n\sigma^2}\sum_{i=1}^{n}\begin{pmatrix}\boldsymbol{x}_i\boldsymbol{x}_i^T & \boldsymbol{x}_i\boldsymbol{z}_i^T\\ \boldsymbol{z}_i\boldsymbol{x}_i^T & \boldsymbol{z}_i\boldsymbol{z}_i^T\end{pmatrix},$$

其中 σ^2 基于全模型来估计; 同理, 基于第 m 个模型的 Fisher 信息阵为

$$J_{S_m|H}=\begin{pmatrix}J_{00|H} & J_{01|H}\Gamma_m^T\\ \Gamma_m J_{10|H} & \Gamma_m J_{11|H}\Gamma_m^T\end{pmatrix};$$

相应的逆矩阵分别记为

$$J_{full|H}^{-1}=\begin{pmatrix}J^{00|H} & J^{01|H}\\ J^{10|H} & J^{11|H}\end{pmatrix}\quad \text{以及}\quad J_{S_m|H}^{-1}=\begin{pmatrix}J^{00,m|H} & J^{01,m|H}\\ J^{10,m|H} & J^{11,m|H}\end{pmatrix}.$$

令

$$\varphi_H=\left\{\left(\frac{\partial\mu}{\partial\boldsymbol{\theta}}\right)^T J_{00|H}^{-1}\frac{\partial\mu}{\partial\boldsymbol{\theta}}+\boldsymbol{\omega}_H^T K_H\boldsymbol{\omega}_H\right\}^{1/2}$$

以及

$$\widehat{\delta}(D_n)=K_H^{1/2}\left\{\sum_{m=1}^{M}w_m(D_n)V_{m|H}\right\}K_H^{-1/2}D_n,$$

其中 $\boldsymbol{\omega}_H=J_{10|H}J_{00|H}^{-1}\frac{\partial\mu}{\partial\boldsymbol{\beta}}-\frac{\partial\mu}{\partial\boldsymbol{\gamma}}$, $K_H=\left(J_{11|H}-J_{10|H}J_{00|H}^{-1}J_{01|H}\right)^{-1}$, $D_n=\widehat{\boldsymbol{\delta}}_{full}=\sqrt{n}(\widehat{\gamma}_{full}-\gamma_0)$, $K_{m|H}=\left(\Gamma_m K_H^{-1}\Gamma_m^T\right)^{-1}$ 以及 $V_{m|H}=K_H^{-1/2}\Gamma_m^T K_{m|H}\Gamma_m K_H^{-1/2}$.

下面, 我们给出一些引理和定理.

引理1. *在固定参数设置 (2.2) 下, 有*

$$\sqrt{n}\begin{pmatrix}\widehat{\boldsymbol{\beta}}_{S_m}-\boldsymbol{\beta}_0\\ \widehat{\gamma}_{S_m}-\gamma_{0,S_m}\end{pmatrix}-J_{S_m|H}^{-1}\begin{pmatrix}J_{01|H}\\ \Gamma_m J_{11|H}\end{pmatrix}\sqrt{n}\boldsymbol{\delta}'\sim\mathcal{N}\left(0,J_{S_m|H}^{-1}\right).$$

进一步, 我们可以得出如下结论.

引理2. *设 $\partial^2\mu(\boldsymbol{\theta})/\left(\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T\right)$, $\partial^2\mu(\boldsymbol{\theta})/\left(\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}^T\right)$ 以及 $\partial^2\mu(\boldsymbol{\theta})/\left(\partial\boldsymbol{\gamma}\partial\boldsymbol{\beta}^T\right)$ 的每个元素均有界, 当 μ 是参数 $\boldsymbol{\gamma}$ 的线性函数时, 在第 m 个模型下, 我们有*

$$\sqrt{n}\left[\{\widehat{\mu}_m-\mu(\boldsymbol{\theta}_0)\}-\boldsymbol{\omega}_H^T\left(\boldsymbol{I}_{q_2}-K_H^{1/2}V_{m|H}K_H^{-1/2}\right)\boldsymbol{\delta}'\right]$$

$$= \left\{\frac{\partial\mu(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\right\}^T J_{00|H}^{-1}M_n - \boldsymbol{\omega}_H^T K_H^{1/2}V_{m|H}K_H^{-1/2}W_n + o_p(1),$$

其中 $M_n \sim \mathcal{N}\left(\mathbf{0}, J_{00|H}\right)$ 且 $W_n \sim \mathcal{N}(\mathbf{0}, K_H)$.

定理1. 在引理 2 的假设下, 模型平均估计 (2.3) 的分布有如下形式

$$\begin{aligned}&\sqrt{n}\left[\{\widehat{\mu}(\boldsymbol{w}(D_n)) - \mu(\boldsymbol{\theta}_0)\} - \boldsymbol{\omega}_H^T\left\{\boldsymbol{\delta}' - \widehat{\delta}(\boldsymbol{\delta}')\right\}\right]\\ = &\left\{\frac{\partial\mu(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\right\}^T J_{00}^{-1}M_n - \boldsymbol{\omega}_H^T\widehat{\delta}(W_n) + o_p(1) \doteq \Lambda + o_p(1),\end{aligned}$$

其中 Λ 服从均值 0 的正态分布, 且方差为

$$\Sigma = \left(\frac{\partial\mu(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right)^T J_{00|H}^{-1}\frac{\partial\mu(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \boldsymbol{\omega}_H^T\mathrm{Var}\widehat{\boldsymbol{\delta}}(W_n)\boldsymbol{\omega}_H.$$

在固定参数设置 (2.2) 下, $\mu(\boldsymbol{\theta}_0)$ 的 $1-\alpha$ 置信区间变成

$$\begin{aligned}CI_{\mu,n} = &\left[\widehat{\mu}(\boldsymbol{w}(D_n)) - \boldsymbol{\omega}_H^T\left\{D_n' - \widehat{\delta}(D_n')\right\} - z_{1-\alpha/2}\varphi_H/\sqrt{n},\right.\\ &\left.\widehat{\mu}(\boldsymbol{w}(D_n)) - \boldsymbol{\omega}_H^T\left\{D_n' - \widehat{\delta}(D_n')\right\} + z_{1-\alpha/2}\varphi_H/\sqrt{n}\right]\end{aligned} \tag{4.7}$$

且 $D_n' = \widehat{\boldsymbol{\delta}}'_{full} = \widehat{\boldsymbol{\gamma}}_{full} - \boldsymbol{\gamma}_0$. 定义

$$T_n = \frac{\sqrt{n}\left[\widehat{\mu}(\boldsymbol{w}(D_n)) - \mu(\boldsymbol{\theta}_0) - \boldsymbol{\omega}_H^T\left\{D_n' - \widehat{\delta}(D_n')\right\}\right]}{\varphi_H}.$$

我们有下列结论.

推论1. 在引理 2 的假设下, T_n 依概率收敛于服从标准正态分布的随机变量, 即 (4.7) 的区间覆盖率趋于预设定的名义水平.

根据这一推论, $\mu(\boldsymbol{\theta}_0)$ 的 $1-\alpha$ 置信区间依赖于 $\boldsymbol{\delta}$. 在固定参数设置下, 当我们把 [1] 所提出的置信区间中的 $\boldsymbol{\delta}$ 替换成 $\sqrt{n}\boldsymbol{\delta}'$ 时, 关于参数 $\boldsymbol{\gamma}$ 的线性函数的区间覆盖率可达到 $1-\alpha$.

下一节, 我们应用 Monte Carlo 模拟来进一步验证这些理论结果.

5 模拟研究

考虑线性回归模型

$$y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{\gamma} + e_i, \quad i = 1, \ldots, n,$$

其中 $\boldsymbol{\beta} = (1,1)^T$, $\boldsymbol{\gamma} = (0, 1.5, 0)^T$, $\boldsymbol{x}_i = (1, x_{i,1})^T$, $\boldsymbol{z}_i = (z_{i,1}, z_{i,2}, z_{i,3})^T$, $(x_{i,1}, z_{i,1}, z_{i,2}, z_{i,3})^T \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q})$; 误差项 e_i, $i = 1, \ldots, n$ 相互独立同正态分布 $\mathcal{N}(0, \sigma^2)$ 且 σ 为 0.5, 1 和 1.5. 设 $\boldsymbol{Q}$ 的对角元均为 1 且非对角元为 $\rho^{|i-j|}$, 其中 $\rho = 0.5$ 和 0.8. 取样本量 $n = 10$, 50 和 100. 基于上述设置, 模拟循环 1000 次, 我们研究参数 β_1, γ_2 以及均值 μ 的置信区间.

下面, 我们考虑一些常用的权重选择准则, 包括 S-AIC、S-BIC 以及 MMA. 基于这些准则的结果在表格中分别用 SAIC、SBIC 以及 MMA 来标记.

<u>**S-AIC 和 S-BIC 权重**</u>: 第 m 个候选模型的 AIC 和 BIC 分别为

$$\mathrm{AIC}_m = -2\sum_{i=1}^{n}\log f(y_i,\widehat{\boldsymbol{\theta}}_m|\boldsymbol{h}_i)+2k_m,\quad \mathrm{BIC}_m = -2\sum_{i=1}^{n}\log f(y_i,\widehat{\boldsymbol{\theta}}_m|\boldsymbol{h}_i)+k_m\log n.$$

从而, 定义 S-AIC 和 S-BIC 权重为

$$\widehat{w}_{\mathrm{xIC},m}=\exp(-\mathrm{xIC}_m/2)/\sum_{m=1}^{M}\exp(-\mathrm{xIC}_m/2),\quad m=1,\dots,M,$$

其中 xIC_m 是第 m 个子模型的 AIC 或 BIC 权值.

<u>**MMA 权重**</u>: 由 [14] 提出的 MMA 权重可通过极小化下式

$$C_n(\boldsymbol{w})=\sum_{i=1}^{n}\left\{y_i-\sum_{m=1}^{M}w_m\widehat{\mu}_m\right\}^2+2\sigma^2\sum_{m=1}^{M}w_mk_m$$

得到, 其中权重向量 $\boldsymbol{w}\equiv(w_1,\dots,w_M)^T$ 且限制在如下单纯形集合 $\mathcal{W}\equiv\{\boldsymbol{w}\in[0,1]^M: 0\le w_k\le 1\}$ 中; σ^2 未知但可基于全模型来估计.

当 $(x_{i,1},z_{i,1},z_{i,2},z_{i,3})^T$ 来自不同分布时, 比如均匀分布、指数分布以及一些混合分布, 我们发现变换协变量的分布对模拟结果的影响甚微. 所以, 这里我们只展示正态分布情形的模拟结果.

基于上述权重, 在表 1-2 中, 我们给出了置信区间的覆盖率和平均长度. 可以看到所有方法的区间覆盖率都达到了预设定的名义水平. 值得注意的是在不同权重准则下, 区间的覆盖率和平均长度是一样的, 这说明 [1] 的置信区间构造方法不受权重影响, 只需权重之和为 1. 这些模拟结果支持了我们的结论.

6 总 结

现有关于模型平均渐近分布理论的研究成果基本上是基于局部误设定的假设, 参数与样本量有关且随样本量增大而变得非常小. 鉴于此假设的非真实性, 我们希望去除这种假设, 研究在一般固定参数设定下, 模型平均估计的统计推断. 本文在一般固定参数设置下, 对于线性回归模型且误差项服从正态分布, 我们证明了 [1] 中的置信区间构造方法对不确定参数的线性函数仍然有效. 换句话说, 尽管 [1] 的渐近分布理论在一般参数设定下无法证实, 但是在上述特殊情况下, 不确定参数的线性函数的置信区间覆盖率可达到预设定的名义水平.

在一般固定参数设置下, 研究模型平均估计量的统计推断问题意义非凡. 未来我们可以研究, 在线性回归模型下, 当误差项来自于其他分布时, [1] 中的置信区间构造方法. 也可以推导基于其他权重选择准则的模型平均估计的渐近分布理论.

参 考 文 献

[1] Hjort N L, Claeskens G. Frequentist model average estimators. *Journal of the American Statistical Association*, 2003, **98**(464): 879–899.

[2] Raftery A E, Zheng Y. Discussion: Performance of Bayesian Model Averaging. *Journal of the American Statistical Association*, 2003, **98**(1): 931-938.

[3] Hoeting J A, Madigan D, Raftery A E, Volinsky C T. Bayesian model averaging: a tutorial. *Statistical Science*, 1999, **14**(4): 382–401.

[4] Claeskens G, Hjort N L. Model Selection and Model Averaging. Cambridge: Cambridge University Press, 2008.

[5] Liu C A. Distribution theory of the least squares averaging estimator. *Journal of Econometrics*, 2015, **186**(1): 142–159.

[6] Leung G, Barron A R. Information theory and mixing least-squares regressions. *IEEE Transactions on Information Theory*, 2006, **52**(8): 3396–3410.

[7] Pötscher, Benedikt M. The Distribution of Model Averaging Estimators and an Impossibility Result regarding Its Estimation. *Lecture Notes-Monograph Series*, 2006, **52**: 113-129.

[8] Hansen B E. Averaging Estimators for Regressions with a Possible Structural Break. *Econometric Theory*, 2009, **25**(6): 1498-1514..

[9] Hansen B E. Averaging estimators for autoregressions with a near unit root. *Journal of Econometrics*, 2010, **158**(1): 142–155.

[10] 朱容, 邹国华. 半参数模型平均估计的渐近理论. 中国科学:数学, 2018, **v.48**(8): 31019–1052. (Zhu R, Zou G H. The asymptotic theory for model averaging in general semiparametric models (in Chinese). *Sci Sin Math*, 2018, **48**(8): 1019–1052)

[11] Zhang X Y, Liu C A. Inference after Model Averaging in Linear Regression Models. *Social Science Electronic Publishing*, 2017.

[12] Mitra P, Lian H, Mitra R,Liang H, Xie M G. A general framework for frequentist model averaging. *Science China Mathematics*, 2019, **62**(2): 205–226.

[13] Ferguson T S. A Course in Large Sample Theory. London: Chapman and Hall, 1996.

[14] Hansen B E. Least squares model averaging. *Econometrica*, 2007, **75**(4): 1175–1189.

附录

引理1 的证明

在线性回归模型下, 当误差项服从正态分布时, 我们有

$$\begin{pmatrix} \frac{\partial \log f(y_i,\boldsymbol{\theta}_0|\boldsymbol{h}_i)}{\partial \boldsymbol{\beta}} \\ \frac{\partial \log f(y_i,\boldsymbol{\theta}_0|\boldsymbol{h}_i)}{\partial \boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sigma^2}\left(-y_i\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{\beta}_0\boldsymbol{x}_i+\boldsymbol{z}_i^T(\boldsymbol{\gamma}_0+\boldsymbol{\delta}')\boldsymbol{x}_i\right) \\ -\frac{1}{\sigma^2}\left(-y_i\boldsymbol{z}_i+\boldsymbol{z}_i^T(\boldsymbol{\gamma}_0+\boldsymbol{\delta}')\boldsymbol{z}_i+\boldsymbol{x}_i^T\boldsymbol{\beta}_0\boldsymbol{z}_i\right) \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{1}{\sigma^2}\left(-y_i\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{\beta}_0\boldsymbol{x}_i+\boldsymbol{z}_i^T\boldsymbol{\gamma}_0\boldsymbol{x}_i\right) \\ -\frac{1}{\sigma^2}\left(-y_i\boldsymbol{z}_i+\boldsymbol{z}_i^T\boldsymbol{\gamma}_0\boldsymbol{z}_i+\boldsymbol{x}_i^T\boldsymbol{\beta}_0\boldsymbol{z}_i\right) \end{pmatrix} + \begin{pmatrix} -\frac{1}{\sigma^2}\boldsymbol{z}_i^T\boldsymbol{\delta}'\boldsymbol{x}_i \\ -\frac{1}{\sigma^2}\boldsymbol{z}_i^T\boldsymbol{\delta}'\boldsymbol{z}_i \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial \log f\left(y_i,(\boldsymbol{\beta}_0^T,\boldsymbol{\gamma}_0^T)^T\middle|\boldsymbol{h}_i\right)}{\partial \boldsymbol{\beta}} \\ \frac{\partial \log f\left(y_i,(\boldsymbol{\beta}_0^T,\boldsymbol{\gamma}_0^T)^T\middle|\boldsymbol{h}_i\right)}{\partial \boldsymbol{\gamma}} \end{pmatrix} + \begin{pmatrix} -\frac{1}{\sigma^2}\boldsymbol{x}_i\boldsymbol{z}_i^T \\ -\frac{1}{\sigma^2}\boldsymbol{z}_i\boldsymbol{z}_i^T \end{pmatrix}\boldsymbol{\delta}'$$

以及

$$\frac{\partial^2 \log f(y_i,\boldsymbol{\theta}_0|\boldsymbol{h}_i)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} = -\frac{1}{\sigma^2}\begin{pmatrix} \boldsymbol{x}_i\boldsymbol{x}_i^T & \boldsymbol{x}_i\boldsymbol{z}_i^T \\ \boldsymbol{z}_i\boldsymbol{x}_i^T & \boldsymbol{z}_i\boldsymbol{z}_i^T \end{pmatrix} = \frac{\partial^2 \log f\left(y_i,(\boldsymbol{\beta}_0^T,\boldsymbol{\gamma}_0^T)^T\middle|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}. \tag{6.8}$$

由 $\sum_{i=1}^n \partial \log f\left(y_i,\left(\widehat{\boldsymbol{\theta}}_{S_m}^T,\boldsymbol{\gamma}_{0,S_m^c}^T\right)^T\middle|\boldsymbol{h}_i\right)\Big/\partial\boldsymbol{\theta}_{S_m}=0$, 将

$$\sum_{i=1}^n \frac{\partial \log f\left(y_i,\left(\widehat{\boldsymbol{\theta}}_{S_m}^T,\boldsymbol{\gamma}_{0,S_m^c}^T\right)^T\middle|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}}$$

在 $\boldsymbol{\theta}_0$ 处 Taylor 展开, 得

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\Bigg[\frac{\partial \log f(y_i,\boldsymbol{\theta}_0|\boldsymbol{h}_i)}{\partial\boldsymbol{\theta}_{S_m}} \\ &\quad+\left(\frac{\partial^2 \log f\left(y_i,(\check{\boldsymbol{\beta}}_0^T,\check{\boldsymbol{\gamma}}_0^T)^T\middle|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T},\frac{\partial^2 \log f\left(y_i,(\check{\boldsymbol{\beta}}_0^T,\check{\boldsymbol{\gamma}}_0^T)^T\middle|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\gamma}_{S_m^c}^T}\right) \\ &\quad\cdot\left(\left(\widehat{\boldsymbol{\beta}}_{S_m}-\boldsymbol{\beta}_0\right)^T,(\widehat{\boldsymbol{\gamma}}_{S_m}-\boldsymbol{\gamma}_{0,S_m})^T,\boldsymbol{0}^T\right)^T+\frac{\partial^2 \log f\left(y_i,(\check{\boldsymbol{\beta}}_0^T,\check{\boldsymbol{\gamma}}_0^T)^T\middle|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\gamma}^T}\boldsymbol{\delta}'\Bigg] \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\Bigg[\frac{\partial \log f(y_i,\boldsymbol{\theta}_0|\boldsymbol{h}_i)}{\partial\boldsymbol{\theta}_{S_m}}+\frac{\partial^2 \log f\left(y_i,(\check{\boldsymbol{\beta}}_0^T,\check{\boldsymbol{\gamma}}_0^T)^T\middle|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T} \\ &\quad\cdot\left(\left(\widehat{\boldsymbol{\beta}}_{S_m}-\boldsymbol{\beta}_0\right)^T,(\widehat{\boldsymbol{\gamma}}_{S_m}-\boldsymbol{\gamma}_{0,S_m})^T\right)^T+\frac{\partial^2 \log f\left(y_i,(\check{\boldsymbol{\beta}}_0^T,\check{\boldsymbol{\gamma}}_0^T)^T\middle|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\gamma}^T}\boldsymbol{\delta}'\Bigg], \end{aligned}$$

其中 $(\check{\boldsymbol{\beta}}_0^T,\check{\boldsymbol{\gamma}}_0^T)^T$ 位于 $\boldsymbol{\theta}_0$ 和 $\left(\widehat{\boldsymbol{\theta}}_{S_m}^T,\boldsymbol{\gamma}_{0,S_m^c}^T\right)^T$ 之间. 因此

$$\begin{aligned} &\begin{pmatrix} \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{S_m}-\boldsymbol{\beta}_0\right) \\ \sqrt{n}\left(\widehat{\boldsymbol{\gamma}}_{S_m}-\boldsymbol{\gamma}_{0,S_m}\right) \end{pmatrix} \\ &= \left(-\frac{1}{n}\sum_{i=1}^n\frac{\partial^2 \log f\left(y_i,(\check{\boldsymbol{\beta}}_0^T,\check{\boldsymbol{\gamma}}_0^T)^T\middle|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T}\right)^{-1} \end{aligned}$$

$$
\begin{aligned}
&\cdot\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f\left(y_i,\boldsymbol{\theta}_0|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}}+\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f\left(y_i,(\check{\boldsymbol{\beta}}_0^T,\check{\boldsymbol{\gamma}}_0^T)^T|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\gamma}^T}\boldsymbol{\delta}'\right]\\
=&\left(-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f\left(y_i,(\check{\boldsymbol{\beta}}_0^T,\check{\boldsymbol{\gamma}}_0^T)^T|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T}\right)^{-1}\\
&\cdot\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\Gamma_m\frac{\partial \log f\left(y_i,\boldsymbol{\theta}_0|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}}+\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f\left(y_i,(\check{\boldsymbol{\beta}}_0^T,\check{\boldsymbol{\gamma}}_0^T)^T|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\gamma}^T}\boldsymbol{\delta}'\right].
\end{aligned}
\tag{6.9}
$$

再由 (6.8), 可得

$$
\begin{aligned}
&\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f\left(y_i,\boldsymbol{\theta}_0|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}}\right\}=\sqrt{n}\begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n}\partial\log f\left(y_i,\boldsymbol{\theta}_0|\boldsymbol{h}_i\right)/\partial\boldsymbol{\beta}\\ \frac{1}{n}\sum_{i=1}^{n}\partial\log f\left(y_i,\boldsymbol{\theta}_0|\boldsymbol{h}_i\right)/\partial\boldsymbol{\gamma}\end{pmatrix}\\
\doteq&\begin{pmatrix}M_n\\ N_n\end{pmatrix}=\frac{1}{\sigma^2}\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}e_i\begin{pmatrix}\boldsymbol{x}_i\\ \boldsymbol{z}_i\end{pmatrix}\sim\mathcal{N}(0,J_{full|H}),
\end{aligned}
$$

$$
-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f\left(y_i,(\check{\boldsymbol{\beta}}_0^T,\check{\boldsymbol{\gamma}}_0^T)^T|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T}=-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f\left(y_i,(\boldsymbol{\beta}_0^T,\boldsymbol{\gamma}_0^T)^T|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\theta}_{S_m}^T}=J_{S_m|H},
\tag{6.10}
$$

$$
\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f\left(y_i,(\check{\boldsymbol{\beta}}_0^T,\check{\boldsymbol{\gamma}}_0^T)^T|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}\partial\boldsymbol{\gamma}^T}\boldsymbol{\delta}'=\begin{pmatrix}-\frac{1}{\sigma^2}\boldsymbol{x}_i\boldsymbol{z}_i^T\\ -\frac{1}{\sigma^2}\Gamma_m\boldsymbol{z}_i\boldsymbol{z}_i^T\end{pmatrix}\boldsymbol{\delta}'=\begin{pmatrix}J_{01|H}\\ \Gamma_m J_{11|H}\end{pmatrix}\boldsymbol{\delta}'
\tag{6.11}
$$

以及

$$
\begin{aligned}
&\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f\left(y_i,\boldsymbol{\theta}_0|\boldsymbol{h}_i\right)}{\partial\boldsymbol{\theta}_{S_m}}\right\}=\sqrt{n}\begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n}\partial\log f\left(y_i,\boldsymbol{\theta}_0|\boldsymbol{h}_i\right)/\partial\boldsymbol{\beta}\\ \frac{1}{n}\sum_{i=1}^{n}\partial\log f\left(y_i,\boldsymbol{\theta}_0|\boldsymbol{h}_i\right)/\partial\boldsymbol{\gamma}_{S_m}\end{pmatrix}\\
\doteq&\begin{pmatrix}M_n\\ N_n^m\end{pmatrix}\sim\mathcal{N}\left(0,J_{S_m|H}\right).
\end{aligned}
\tag{6.12}
$$

因此, 将 (6.10)-(6.12) 代入 (6.9), 可得

$$
\sqrt{n}\begin{pmatrix}\widehat{\boldsymbol{\beta}}_{S_m}-\boldsymbol{\beta}_0\\ \widehat{\boldsymbol{\gamma}}_{S_m}-\boldsymbol{\gamma}_{0,S_m}\end{pmatrix}=J_{S_m|H}^{-1}\left\{\begin{pmatrix}J_{01|H}\\ \Gamma_m J_{11|H}\end{pmatrix}\sqrt{n}\boldsymbol{\delta}'+\begin{pmatrix}M_n\\ N_n^m\end{pmatrix}\right\}
\tag{6.13}
$$

以及

$$
\sqrt{n}\begin{pmatrix}\widehat{\boldsymbol{\beta}}_{S_m}-\boldsymbol{\beta}_0\\ \widehat{\boldsymbol{\gamma}}_{S_m}-\boldsymbol{\gamma}_{0,S_m}\end{pmatrix}-J_{S_m|H}^{-1}\begin{pmatrix}J_{01|H}\\ \Gamma_m J_{11|H}\end{pmatrix}\sqrt{n}\boldsymbol{\delta}'\sim\mathcal{N}\left(0,J_{S_m|H}^{-1}\right).
\tag{6.14}
$$

引理 2 的证明

记 $\boldsymbol{\delta}'_{S_m}=\left(\delta'_{i_1},\cdots,\delta'_{i_{k_m-q_1}}\right)^T$ 和 $\boldsymbol{\delta}'_{S_m^c}=\left(\delta'_{i_{k_m-q_1+1}},\cdots,\delta'_{i_{q_2}}\right)^T$. 将 $\widehat{\mu}_m$ 在 $\boldsymbol{\theta}_0$ 处 Taylor 展开得

$$
\widehat{\mu}_m-\mu\left(\boldsymbol{\theta}_0\right)=\left\{\frac{\partial\mu\left(\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\beta}}\right\}^T\left(\widehat{\boldsymbol{\beta}}_{S_m}-\boldsymbol{\beta}_0\right)+\left\{\frac{\partial\mu\left(\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\gamma}_{S_m}}\right\}^T\left(\widehat{\boldsymbol{\gamma}}_{S_m}-\boldsymbol{\gamma}_{0,S_m}\right)-\left\{\frac{\partial\mu\left(\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\gamma}}\right\}^T\boldsymbol{\delta}'
$$

$$+\begin{pmatrix}\widehat{\boldsymbol\beta}_{S_m}-\boldsymbol\beta_0\\ \widehat{\boldsymbol\gamma}_{S_m}-\boldsymbol\gamma_{0,S_m}-\boldsymbol\delta'_{S_m}\\ -\boldsymbol\delta'_{S_m^c}\end{pmatrix}^T \Pi_m\begin{pmatrix}\frac{\partial^2\mu(\overline{\boldsymbol\theta}_0)}{\partial\boldsymbol\beta\partial\boldsymbol\beta^T} & \frac{\partial^2\mu(\overline{\boldsymbol\theta}_0)}{\partial\boldsymbol\beta\partial\boldsymbol\gamma^T}\\ \frac{\partial^2\mu(\overline{\boldsymbol\theta}_0)}{\partial\boldsymbol\gamma\partial\boldsymbol\beta^T} & \frac{\partial^2\mu(\overline{\boldsymbol\theta}_0)}{\partial\boldsymbol\gamma\partial\boldsymbol\gamma^T}\end{pmatrix}\Pi_m^T\begin{pmatrix}\widehat{\boldsymbol\beta}_{S_m}-\boldsymbol\beta_0\\ \widehat{\boldsymbol\gamma}_{S_m}-\boldsymbol\gamma_{0,S_m}-\boldsymbol\delta'_{S_m}\\ -\boldsymbol\delta'_{S_m^c}\end{pmatrix}, \tag{6.15}$$

其中 $\overline{\boldsymbol\theta}_0$ 位于 $\boldsymbol\theta_0$ 和 $\left(\widehat{\boldsymbol\theta}_{S_m}^T, \widehat{\boldsymbol\gamma}_{S_m}^T, \boldsymbol\gamma_{0,S_m^c}^T\right)^T$ 之间. 类似于 (6.14) 的证明, 可得

$$\sqrt{n}\begin{pmatrix}\widehat{\boldsymbol\beta}_{S_m}-\boldsymbol\beta_0\\ \widehat{\boldsymbol\gamma}_{S_m}-\boldsymbol\gamma_{0,S_m}-\boldsymbol\delta'_{S_m}\end{pmatrix}\sim\mathcal{N}\left(0, J_{S_m|H}^{-1}\right). \tag{6.16}$$

当 μ 是参数 $\boldsymbol\gamma$ 的线性函数时, $\frac{\partial^2\mu(\overline{\boldsymbol\theta}_0)}{\partial\boldsymbol\gamma\partial\boldsymbol\gamma^T}=0$. 又 $\frac{\partial^2\mu(\overline{\boldsymbol\theta}_0)}{\partial\boldsymbol\beta\partial\boldsymbol\beta^T}$, $\frac{\partial^2\mu(\overline{\boldsymbol\theta}_0)}{\partial\boldsymbol\beta\partial\boldsymbol\gamma^T}$ 以及 $\frac{\partial^2\mu(\overline{\boldsymbol\theta}_0)}{\partial\boldsymbol\gamma\partial\boldsymbol\beta^T}$ 有界, 结合 (6.16), 得

$$\begin{pmatrix}\widehat{\boldsymbol\beta}_{S_m}-\boldsymbol\beta_0\\ \widehat{\boldsymbol\gamma}_{S_m}-\boldsymbol\gamma_{0,S_m}-\boldsymbol\delta'_{S_m}\\ -\boldsymbol\delta'_{S_m^c}\end{pmatrix}^T \Pi_m\begin{pmatrix}\frac{\partial^2\mu(\overline{\boldsymbol\theta}_0)}{\partial\boldsymbol\beta\partial\boldsymbol\beta^T} & \frac{\partial^2\mu(\overline{\boldsymbol\theta}_0)}{\partial\boldsymbol\beta\partial\boldsymbol\gamma^T}\\ \frac{\partial^2\mu(\overline{\boldsymbol\theta}_0)}{\partial\boldsymbol\gamma\partial\boldsymbol\beta^T} & \frac{\partial^2\mu(\overline{\boldsymbol\theta}_0)}{\partial\boldsymbol\gamma\partial\boldsymbol\gamma^T}\end{pmatrix}\Pi_m^T\begin{pmatrix}\widehat{\boldsymbol\beta}_{S_m}-\boldsymbol\beta_0\\ \widehat{\boldsymbol\gamma}_{S_m}-\boldsymbol\gamma_{0,S_m}-\boldsymbol\delta'_{S_m}\\ -\boldsymbol\delta'_{S_m^c}\end{pmatrix}=O_p\left(\frac{1}{n}\right).$$

将 (6.13) 代入 (6.15), 可得

$$\begin{aligned}
&\sqrt{n}\left[\{\widehat{\mu}_m-\mu(\boldsymbol\theta_0)\}+\left\{\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\gamma}\right\}^T\boldsymbol\delta'-\begin{pmatrix}\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\beta}\\ \frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\gamma_{S_m}}\end{pmatrix}^T J_{S_m|H}^{-1}\begin{pmatrix}J_{01|H}\\ \Gamma_m J_{11|H}\end{pmatrix}\boldsymbol\delta'\right]\\
=&\begin{pmatrix}\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\beta}\\ \frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\gamma_{S_m}}\end{pmatrix}^T\left[\begin{pmatrix}\sqrt{n}\left(\widehat{\boldsymbol\beta}_{S_m}-\boldsymbol\beta_0\right)\\ \sqrt{n}\left(\widehat{\boldsymbol\gamma}_{S_m}-\boldsymbol\gamma_{0,S_m}\right)\end{pmatrix}-J_{S_m|H}^{-1}\begin{pmatrix}J_{01}\\ \Gamma_m J_{11|H}\end{pmatrix}\sqrt{n}\boldsymbol\delta'\right]\\
=&\begin{pmatrix}\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\beta}\\ \frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\gamma_{S_m}}\end{pmatrix}^T\mathbf{J}_{S_m|H}^{-1}\begin{pmatrix}M_n\\ N_n^m\end{pmatrix}+o_p(1)\doteq\Lambda_m+o_p(1).
\end{aligned}$$

令 $W_n=J^{10|H}M_n+J^{11|H}N_n$. 由 [1] 中引理 3.3 的证明, 同理可得

$$\begin{aligned}
&\begin{pmatrix}\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\beta}\\ \frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\gamma_{S_m}}\end{pmatrix}^T J_{S_m|H}^{-1}\begin{pmatrix}J_{01|H}\\ \Gamma_m J_{11|H}\end{pmatrix}\boldsymbol\delta'-\left\{\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\gamma}\right\}^T\boldsymbol\delta'\\
=&\begin{pmatrix}\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\beta}\\ \frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\gamma_{S_m}}\end{pmatrix}^T\begin{pmatrix}J^{00,m|H} & J^{01,m|H}\\ J^{10,m|H} & J^{11,m|H}\end{pmatrix}\begin{pmatrix}J_{01|H}\\ \Gamma_m J_{11|H}\end{pmatrix}\boldsymbol\delta'-\left\{\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\gamma}\right\}^T\boldsymbol\delta'\\
=&\left\{\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\beta}\right\}^T\left(J^{00,m}J_{01|H}+J^{01,m|H}\Gamma_m J_{11|H}\right)\boldsymbol\delta'\\
&+\left\{\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\gamma_{S_m}}\right\}^T\left(J^{10,m|H}J_{01|H}+J^{11,m|H}\Gamma_m J_{11|H}\right)\boldsymbol\delta'-\left\{\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\gamma}\right\}^T\boldsymbol\delta'\\
=&\left\{\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\beta}\right\}^T J_{00|H}^{-1}J_{01|H}\left(\boldsymbol I_{q_2}-K_H^{1/2}V_m K_H^{-1/2}\right)\boldsymbol\delta'\\
&+\left\{\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\gamma}\right\}^T\left(\Gamma_m^T K_{m|H}\Gamma_m K_H^{-1}\right)\boldsymbol\delta'-\left\{\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\gamma}\right\}^T\boldsymbol\delta'\\
=&\boldsymbol\omega_H^T\left(\boldsymbol I_{q_2}-K_H^{1/2}V_m K_H^{-1/2}\right)\boldsymbol\delta'
\end{aligned}$$

以及

$$\Lambda_m=\begin{pmatrix}\frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\beta}\\ \frac{\partial\mu(\boldsymbol\theta_0)}{\partial\boldsymbol\gamma_{S_m|H}}\end{pmatrix}^T J_{S_m|H}^{-1}\begin{pmatrix}M_n\\ N_n^m\end{pmatrix}$$

$$
\begin{aligned}
&= \begin{pmatrix} \frac{\partial\mu(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}} \\ \frac{\partial\mu(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\gamma}_{S_m}} \end{pmatrix}^T \begin{pmatrix} J^{00,m|H} & J^{01,m|H} \\ J^{10,m|H} & J^{11,m|H} \end{pmatrix} \begin{pmatrix} M_n \\ N_n^m \end{pmatrix} \\
&= \left\{\frac{\partial\mu(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\right\}^T \left(J^{00,m|H}M_n + J^{01,m|H}N_n^m\right) + \left\{\frac{\partial\mu(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\gamma}_{S_m}}\right\}^T \left(J^{10,m|H}M_n + J^{11,m|H}N_n^m\right) \\
&= \left\{\frac{\partial\mu(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\right\}^T \left(J_{00|H}^{-1}M_n - J_{00|H}^{-1}J_{01|H}K_H^{1/2}V_mK_H^{-1/2}W_n\right) + \left\{\frac{\partial\mu(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\gamma}_{S_m}}\right\}^T K_{m|H}\Gamma_mK_H^{-1}W_n \\
&= \left\{\frac{\partial\mu(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\right\}^T J_{00|H}^{-1}M_n - \boldsymbol{\omega}_H^TK_H^{1/2}V_mK_H^{-1/2}W_n.
\end{aligned}
$$

因此, 我们可得到类似于[1] 中引理3.3 的结论：

$$
\begin{aligned}
&\sqrt{n}\left[\{\widehat{\mu}_m - \mu(\boldsymbol{\theta}_0)\} - \boldsymbol{\omega}_H^T\left(\boldsymbol{I}_{q_2} - K_H^{1/2}V_mK_H^{-1/2}\right)\boldsymbol{\delta}'\right] \\
&= \left\{\frac{\partial\mu(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\right\}^T J_{00|H}^{-1}M_n - \boldsymbol{\omega}_H^TK_H^{1/2}V_mK_H^{-1/2}W_n + o_p(1).
\end{aligned}
$$

定理 1 的证明

令 $\boldsymbol{I}_{q_2}$ 为 q_2 维单位矩阵. 仿照 (6.9) 的证明可得

$$
\begin{aligned}
D_n &= \sqrt{n}(\widehat{\boldsymbol{\gamma}}_{full} - \boldsymbol{\gamma}_0) \\
&= \begin{pmatrix} \boldsymbol{0}_{q_2\times q_1} & \boldsymbol{I}_{q_2} \end{pmatrix} J_{full|H}^{-1}\sqrt{n}\left\{\begin{pmatrix} J_{01|H} \\ J_{11|H} \end{pmatrix}\boldsymbol{\delta}' + \begin{pmatrix} M_n \\ N_n \end{pmatrix}\right\}
\end{aligned}
$$

以及

$$
D_n - \sqrt{n}\boldsymbol{\delta}' = J^{10|H}M_n + J^{11|H}N_n = W_n \sim \mathcal{N}(0, K_H).
$$

注意随机向量 W_n 和 M_n 是独立的. 由于对所有候选模型 m, 随机变量 $w_m(D_n)$ 和 $\widehat{\mu}_m$ 可表示成同一随机向量 $\left(M_n^T, N_n^T\right)^T$ 的函数, 并且 $w_m(D_n)$ 是 D_n 的连续函数, 从而

$$
\begin{aligned}
&\sqrt{n}\left(\left\{\sum_{m=1}^M w_m(D_n)\widehat{\mu}_m - \mu(\boldsymbol{\theta}_0)\right\} - \boldsymbol{\omega}_H^T\left[\boldsymbol{I}_{q_2} - K^{1/2}\left\{\sum_{m=1}^M w_m(D_n)H_m\right\}K^{-1/2}\right]\boldsymbol{\delta}'\right) \\
&= \sqrt{n}\left[\left\{\sum_{m=1}^M w_m(D_n)\widehat{\mu}_m - \mu(\boldsymbol{\theta}_0)\right\} - \boldsymbol{\omega}_H^T\left\{\boldsymbol{\delta}' - \widehat{\delta}(\boldsymbol{\delta}')\right\}\right] \\
&= \left\{\frac{\partial\mu(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\right\}^T J_{00|H}^{-1}M_n - \boldsymbol{\omega}_H^T\widehat{\boldsymbol{\delta}}(W_n) + o_p(1) \doteq \Lambda + o_p(1),
\end{aligned}
$$

其中 $\widehat{\boldsymbol{\delta}}(W_n) = K^{1/2}\left\{\sum_{m=1}^M w_m(W_n)H_m\right\}K^{-1/2}W_n$, 此外, 无论是在固定参数设置还是在局部误设定假设下, Λ 服从均值为 0 且方差为 Σ 的正态分布, 其中

$$
\Sigma = \left(\frac{\partial\mu}{\partial\boldsymbol{\theta}}\right)^T J_{00|H}^{-1}\frac{\partial\mu}{\partial\boldsymbol{\theta}} + \boldsymbol{\omega}_H^T\mathrm{Var}\widehat{\boldsymbol{\delta}}(W_n)\boldsymbol{\omega}_H.
$$

推论 1 的证明

由定理 1 的证明, 可得

$$
\begin{aligned}
T_n &= \frac{\sqrt{n}\left[\widehat{\mu}(\boldsymbol{w}(D_n)) - \mu(\boldsymbol{\theta}_0) - \boldsymbol{\omega}_H^T\left\{D_n' - \widehat{\delta}(D_n')\right\}\right]}{\varphi_H} \\
&= \frac{\sqrt{n}\left[\widehat{\mu}(\boldsymbol{w}(D_n)) - \mu(\boldsymbol{\theta}_0)\right] - \boldsymbol{\omega}_H^T\left\{D_n - \widehat{\delta}(D_n)\right\}}{\varphi_H}
\end{aligned}
$$

$$= \frac{\sqrt{n}\left[\widehat{\mu}\left(\boldsymbol{w}\left(D_n\right)\right)-\mu\left(\boldsymbol{\theta}_0\right)-\boldsymbol{\omega}_H^T\left\{\boldsymbol{\delta}'-\widehat{\delta}\left(\boldsymbol{\delta}'\right)\right\}\right]-\boldsymbol{\omega}_H^T\left(D_n-\sqrt{n}\boldsymbol{\delta}'\right)+\boldsymbol{\omega}_H^T\left\{\widehat{\delta}\left(D_n\right)-\sqrt{n}\widehat{\delta}\left(\boldsymbol{\delta}'\right)\right\}}{\varphi_H}$$

$$= \frac{\left\{\frac{\partial\mu(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\right\}^T J_{00|H}^{-1} M_n - \boldsymbol{\omega}_H^T W_n}{\varphi_H} + o_p(1) = Z_H + o_p(1),$$

其中 $Z_H \sim \mathcal{N}(0,1)$.

表 1 在线性回归模型下, 当 $\rho = 0.5$ 时, 95% 置信区间的覆盖率 (CP(95)) 和长度 (Len(95))

(Table 1 Coverage Probability and Length of 95% confidence intervals (CP(95) and Len(95)): $\rho = 0.5$)

			$\sigma = 0.5$			$\sigma = 1$			$\sigma = 1.5$		
n		方法	β_1	γ_2	μ	β_1	γ_2	μ	β_1	γ_2	μ
50	CP(95)	SAIC	0.950	0.947	0.949	0.939	0.953	0.950	0.957	0.945	0.945
		SBIC	0.950	0.947	0.949	0.939	0.953	0.950	0.957	0.945	0.945
		MMA	0.950	0.947	0.949	0.939	0.953	0.950	0.957	0.945	0.945
	Len(95)	SAIC	0.3413	0.3809	0.6319	0.6828	0.7676	1.2721	1.0259	1.1409	1.9238
		SBIC	0.3413	0.3809	0.6319	0.6828	0.7676	1.2721	1.0259	1.1409	1.9238
		MMA	0.3413	0.3809	0.6319	0.6828	0.7676	1.2721	1.0259	1.1409	1.9238
100	CP(95)	SAIC	0.945	0.958	0.954	0.956	0.951	0.958	0.959	0.941	0.936
		SBIC	0.945	0.958	0.954	0.956	0.951	0.958	0.959	0.941	0.936
		MMA	0.945	0.958	0.954	0.956	0.951	0.958	0.959	0.941	0.936
	Len(95)	SAIC	0.2322	0.2604	0.4350	0.4673	0.5233	0.8649	0.7008	0.7819	1.3102
		SBIC	0.2322	0.2604	0.4350	0.4673	0.5233	0.8649	0.7008	0.7819	1.3102
		MMA	0.2322	0.2604	0.4350	0.4673	0.5233	0.8649	0.7008	0.7819	1.3102
200	CP(95)	SAIC	0.956	0.956	0.952	0.951	0.962	0.957	0.952	0.953	0.954
		SBIC	0.956	0.956	0.952	0.951	0.962	0.957	0.952	0.953	0.954
		MMA	0.956	0.956	0.952	0.951	0.962	0.957	0.952	0.953	0.954
	Len(95)	SAIC	0.1622	0.1810	0.3031	0.3258	0.3629	0.6061	0.4867	0.5439	0.8947
		SBIC	0.1622	0.1810	0.3031	0.3258	0.3629	0.6061	0.4867	0.5439	0.8947
		MMA	0.1622	0.1810	0.3031	0.3258	0.3629	0.6061	0.4867	0.5439	0.8947

表 2 在线性回归模型下, 当 $\rho = 0.8$ 时, 95% 置信区间的覆盖率 (CP(95) 和长度 Len(95)

(Table 2 Coverage Probability and Length of 95% confidence intervals (CP(95) and Len(95)): $\rho = 0.8$)

			$\sigma = 0.5$			$\sigma = 1$			$\sigma = 1.5$		
n		方法	β_1	γ_2	μ	β_1	γ_2	μ	β_1	γ_2	μ
50	CP(95)	SAIC	0.954	0.960	0.939	0.937	0.960	0.952	0.945	0.937	0.944
		SBIC	0.954	0.960	0.939	0.937	0.960	0.952	0.945	0.937	0.944
		MMA	0.954	0.960	0.939	0.937	0.960	0.952	0.945	0.937	0.944
	Len(95)	SAIC	0.4933	0.6330	0.6312	0.9886	1.2614	1.2691	1.4685	1.8644	1.8739
		SBIC	0.4933	0.6330	0.6312	0.9886	1.2614	1.2691	1.4685	1.8644	1.8739
		MMA	0.4933	0.6330	0.6312	0.9886	1.2614	1.2691	1.4685	1.8644	1.8739
100	CP(95)	SAIC	0.945	0.942	0.946	0.951	0.952	0.958	0.946	0.942	0.933
		SBIC	0.945	0.942	0.946	0.951	0.952	0.958	0.946	0.942	0.933
		MMA	0.945	0.942	0.946	0.951	0.952	0.958	0.946	0.942	0.933
	Len(95)	SAIC	0.3366	0.4289	0.4342	0.6746	0.8630	0.8672	1.0138	1.2902	1.3169
		SBIC	0.3366	0.4289	0.4342	0.6746	0.8630	0.8672	1.0138	1.2902	1.3169
		MMA	0.3366	0.4289	0.4342	0.6746	0.8630	0.8672	1.0138	1.2902	1.3169
200	CP(95)	SAIC	0.954	0.946	0.943	0.944	0.948	0.943	0.956	0.952	0.950
		SBIC	0.954	0.946	0.943	0.944	0.948	0.943	0.956	0.952	0.950
		MMA	0.954	0.946	0.943	0.944	0.948	0.943	0.956	0.952	0.950
	Len(95)	SAIC	0.2353	0.3006	0.3012	0.4674	0.5982	0.6045	0.7047	0.8973	0.8968
		SBIC	0.2353	0.3006	0.3012	0.4674	0.5982	0.6045	0.7047	0.8973	0.8968
		MMA	0.2353	0.3006	0.3012	0.4674	0.5982	0.6045	0.7047	0.8973	0.8968